Evolutionary Progress and Levels of Selection

John Maynard Smith

The concept of progress has a bad name in evolutionary biology. There are a number of good reasons for this. In the first part of this essay, I review these reasons: in the second part, I suggest a way of classifying the stages through which life has passed. Although biologists may be reluctant to see these stages as representing "advance," they are progressive in one sense: the sequence in which they have occurred is not arbitrary, since each stage was a necessary precondition for the next. But first, what are the reasons for distrusting the idea of progress?

The first is semantic. How shall we define progress? There is an obvious danger of anthropomorphism: the more similar an organism is to us, the more advanced it seems. Historically, the idea of evolutionary progress derives from the medieval ladder of life, although to be fair, it was not assumed that we stood at the top of the ladder. There were angels and archangels above us, as there were beasts below. The snag, of course, is that life is better represented by a branching tree than by a linear progression. Does the claim of progress mean that there has been progress in all the branches, or, more plausibly, in at least one?

There is a natural wish to measure progress in terms of morphological complexity: there must be some sense in which an oak tree or an elephant is structurally more complex than anything that existed, say, 10^9 years ago. However, even attempts to measure morphological difference, without any implication of progress, have proved to be full of pitfalls. When comparing different taxa, it seems that the best one can do is measure the number of cell types present. An easier measure of complexity is the amount of coding DNA present in the genome. Some estimates are given in Table 1. There is some slight tendency for organisms that we think of as structurally more

Table 1

DNA content per haploid genome

	Genome size (pg)	Percent of genome coding for protein	Coding DNA (pg)
Escherichia coli	0.004	100.0	0.4
Saccharomyces cerevisiae	0.009	69.0	0.62
Caenorhabditis elegans	0.088	25.0	2.2
Drosophila melanogaster	0.18	33.0	5.9
Homo sapiens	3.5	20.0	70.0
Triturus cristatus	19.0	3.0	57.0
Protopterus aethiopicus	142.0	0.8	114.0
Arabidopsis thaliana	0.2	31.0	6.2
Fritillaria assyriaca	127.0	0.02	2.5

Data from Cavalier-Smith 1985. 1 pg of DNA corresponds to approximately 10^9 base pairs. The values of the percent DNA coding for protein are the averages of different estimates, and are only approximate.

complex to have more coding DNA, but it is not very striking. In any case, it is well to remember that the genome is best thought of as a set of instructions for making an organism, and not as a description of the adult structure. One form of progress, therefore, could be the evolution of a more efficient (in terms of quantity of information) set of instructions for making the same structure: if so, progress would be accompanied by a decrease in coding DNA.

A second difficulty is that our theory of evolution does not predict an increase in anything. At first sight, Fisher's (1930) "fundamental theorem of natural selection" might seem to predict an increase in "mean fitness," but it would be a mistake to think that there is any quantity that necessarily increases, as entropy increases in a closed physical system. Thus consider the following imaginary experiments. First, a physicist is provided with any information he wants about the state of a closed physical system at two points in time, but is not told which state is the earlier one. Provided that entropy was not at a maximum on both occasions (that is, provided that something interesting had happened), she could say which state is the earlier one. Now consider a closed biological system: by "closed" I mean that no living organisms enter or leave the system. There is nothing in Fisher's theorem that would enable a biologist to say which state was earlier. Thus even for a

single species in a constant physical environment, there can be a continuous cycle of gene frequencies if fitnesses are frequency dependent. Hence, given data on a population at two moments, one could not tell which was earlier. For evolution in multi-species communities, cycles and other more complex dynamics are likely to be the rule. For evolutionary changes, therefore, it is hard to put an arrow on time, let alone demonstrate the inevitability of progress.

Finally, there is little empirical support for the concept of progress. Others will discuss the fossil record from this point of view, so I will consider the matter from the opposite extreme -- the study of evolution in vitro (Orgel 1979). If RNA molecules are replicated in a test tube by the viral enzyme, Qß replicase, it is possible to follow the base sequence and secondary structure of the population of molecules. More or less independently of the starting point (that is, of the base sequence of the RNA molecule used as an initial primer of the system), the end point is a rather small molecule, some 200 bases long, with a particular sequence and structure that enable it to be replicated particularly rapidly. In this simple and well-defined system, natural selection does not lead to continuing change, still less to anything that could be recognized as an increase in complexity: it leads to a stable and rather simple end point. This raises the following simple, and I think unanswered, question: What features must be present in a system if it is to lead to indefinitely continuing evolutionary change?

To summarize, progress is hard to define. If we equate progress to increase in complexity, the best we can do is to measure complexity in terms of the quantity of coding DNA, or, in eukaryotes, by the number of distinct cell types. Theoretically, there is no quantity, analogous to entropy, that necessarily increases in an evolving population. Observationally, natural selection in the simplest systems does not lead to indefinitely continuing change, let alone to progress. It is true that, on the largest scale, evolution does seem to have given rise to increasingly complex organisms. But, since the first living things were necessarily simple, it is not surprising that the most complex things alive today are more complex than their first progenitors.

This is a rather boring conclusion. It is more fruitful to attempt to categorize the levels of complexity through which life has passed. In doing this, I use Darwinian theory to classify levels of complexity, but I do not

Table 2

Levels of complexity, differing in the organization of the genetic material

1. Replicating molecules	5. Multicellular organisms
2. Populations of molecules in compartments	6. Demes; social groups
3. Prokaryotic cells	7. Species
4. Eukaryotic cells	8. Groups with cultural inheritance

claim that the various levels are predictable from that theory. Darwin's theory can be summarized as follows: Any population of entities with the properties of multiplication (one entity can give rise to many), variation (entities are not all alike, and some kinds are more likely to survive and multiply than others), and heredity (like begets like) will evolve. Evolution on earth has occurred because living organisms have these properties. A major problem in current evolutionary theory is to identify the relevant entities. For Darwin, they were individual organisms. Today, they have been variously identified as genes, genomes, organisms, demes, species, communities and even, somewhat illogically, as the whole biosphere. In what follows, I shall refer to entities with the three necessary properties as "units of evolution." I prefer this phrase to "units of selection," because selection often acts on objects that lack heredity, and so can not evolve by natural selection. Units of evolution can be expected to evolve characteristics that ensure their own survival and reproduction.

During evolutionary time, the entities that constitute units of evolution have passed through a number of levels of complexity. A somewhat speculative series of stages is shown in Table 2. The table is speculative in two ways. There is no direct evidence of the existence of stages 1 and 2, but we infer that they existed because it is hard to see how life could get started otherwise. There is no doubt that stages 3 to 8 exist, but it is arguable whether the last three should be regarded as units of evolution.

A classification of this kind is, of course, a truism. It becomes interesting if we ask the following questions: (i) What is the nature of the genetic information that is passed from generation to generation at each stage? (ii) How is the integrity of that information protected against selection at lower levels? (iii) How did natural selection bring about the

transition from one stage to another, since, at each transition, selection for "selfishness" between entities at the lower level would tend to counteract the change.

A full answer to these questions would constitute a solution to most of the outstanding problems in evolutionary biology. I can make only a few brief suggestions. My main aim is to bring out the analogies between the different transitions. In each case, the difficulty is the same: how is it that selection at a lower level does not disrupt integration at a higher one?

Replicating Molecules

It is widely believed that the first Darwinian entities were RNA-like molecules, replicating by complementary base pairing. Orgel has recently suggested that RNA itself may not be primitive, and that in the first replicators ribose was replaced by some simpler compound: what is primitive is replication that is dependent on base pairing. In the absence of enzymes, replication accuracy would be low, and this would limit the length of the molecules whose sequence could be maintained by selection to, very approximately, 10-100 bases. This would rule out the evolution of a genetic message long enough to code for the enzymes needed to improve the accuracy of replication. This difficulty led Eigen and Schuster (1979) to suggest the evolution of "hypercycles." Essentially, a hypercycle is a collection of molecules of different kinds, functionally arranged in a cycle, . . . A-B-C-D-A-B . . ., so that the rate of synthesis of each molecule in the cycle is accelerated by the concentration of the molecule immediately preceding it. Such a system can be chemically stable, and can transmit a greater total quantity of information than a single molecule. However, selection will not readily incorporate mutations that increase the efficiency of the cycle as a whole.

Thus suppose that a mutation in molecule A makes it better at stimulating the synthesis of B. This would improve the efficiency of the cycle as a whole, but the mutant would not increase under natural selection. By analogy, a mutation in grass that makes it better for nourishing antelopes would not spread. Selection at the level of individual molecules need not lead to

improvement of the higher level system, the hypercycle. This difficulty can be overcome by enclosing the population of molecules within a "compartment," leading to the next stage.

Populations of Molecules in Compartments

If populations of replicating molecules are enclosed -- in spherical membranes, or in the droplets of a coacervate as suggested by Oparin (1957), or in proteinoid microspheres as suggested by Fox (1984) -- the difficulty is partly overcome. A mutant A molecule that accelerates the synthesis of B will, because of the cyclic arrangement, accelerate its own synthesis, whereas in the absence of enclosure it would accelerate the synthesis of other nonmutant A molecules. However, if the compartments, rather than the individual molecules, are to be the units of evolution, they must possess the properties of multiplication, variation and heredity. They will have heredity, because when a compartment splits, the molecular species it contains will be passed on. But it must also be true that a compartment must grow and divide more rapidly if the molecules it contains replicate faster.

Even with between-compartment selection, however, it is not clear how far the hypercycle is proof against selfish mutations at the molecular level. Szathmary and Demeter (1987) have recently argued that further evolution at the compartment level requires an additional feature, which they refer to as the "stochastic corrector." They require that the numbers of molecules of each kind in a compartment be small. This has the effect that chance differences between compartments arise at division, thus making selection at the higher level more effective. This is a feature of all transitions between levels: we shall meet it again when discussing the origin of multicellular organisms, and the effectiveness of interdemic selection.

Selection between compartments is more effective if the number of molecules is small. However, if molecules are partitioned randomly between daughters, there is a risk of a daughter receiving no molecules of a particular type. This risk can be avoided by joining all the types, which we can now call genes, end to end on a chromosome, of which exactly one copy is passed to each daughter cell. This is possible only when the accuracy of replication

is high enough. Initially, when replication is inaccurate, it must be possible for selection, within compartments, to act separately on each type, to prevent the accumulation of errors, and this requires a population within the cell. We are still some way from understanding these transitions, from replicating molecules to populations of interacting molecules within compartments, and thence to cells with chromosomes.

Prokaryotic Cells

During most of the history of life, the predominant units of evolution have been prokaryotic cells whose genetic information is carried in a single circular chromosome. The theory of prokaryotic evolution is in its infancy. We have to learn the evolutionary significance of bacterial conjugation and transformation, and the roles of phages, plasmids and transposons. In the context of the evolution of increasing complexity, Cavalier-Smith (1981) has argued that the crucial limitation lies in the fact that there can be only one point of initiation of replication per chromosome, if one copy is to be transmitted to each daughter cell. This places an upper limit on the quantity of genetic information per cell.

Eukaryotic Cells with Sex

The symbiotic origin of eukaryotic cells brought together genetic information from several different prokaryotic lineages. A more important factor in permitting an increase in complexity is the existence of several chromosomes, each with multiple sites of initiation of replication, made possible by the evolution of mitosis. Since natural selection lacks foresight, we cannot argue that mitosis evolved because it made possible an increase in complexity. However, a change which speeded up the replication of a given quantity of DNA would also permit a future increase in that quantity.

It is the origin of meiosis and syngamy, however, that raises the real difficulties for Darwinian theory, because the genetic information present in an individual is not passed as a unit to its offspring. This has led Williams (1966)

and Dawkins (1976) to argue that, since the gene is the largest unit that is transmitted without recombination for reasonably long periods, we must see the genes as the unit of selection. I think this is correct (although, for the reasons given above, I would prefer the term "unit of evolution"). For many purposes, however, we can safely think in terms of individuals. The reason is that, almost always, there is no within-individual selection between genes at a locus: in other words, meiotic drive is very rare. Consequently, a gene will increase in frequency only if, on average, it increases the fitness of its carriers. "On average" implies averaging over environments, and over the genes present in the population at other loci. Therefore, although genomes are not replicated as units, we can understand why natural selection leads to adaptation at the individual level.

To achieve individual adaptation, it is not sufficient that selection should, typically, act on individuals: it is also necessary that selection between genes within individuals should be unimportant. The argument, by Mayr (1963) and others, that because the genes in an organism are coadapted there must be something inadequate in "beanbag genetics," is a misunderstanding. Of course the genes in an organism are coadapted, in the sense that a set of rabbit genes or of mouse genes work well together, but a set consisting half of rabbit genes and half of mouse genes would not. This is so because a gene only spreads under selection if its effects are coadapted to the other genes in the population. Coadaptation is a consequence of beanbag genetics. In most species, there is no genetic entity much larger than a gene which is transmitted as a unit for long enough to serve as a unit of evolution.

All this seems tolerably clear. The real difficulty arises when we ask why meiosis and syngamy arose in the first place, and why they are maintained in most eukaryotes. I will return to this problem when discussing the species level.

Multicellular Eukaryotes

Higher animals and plants differ from protists in that there are, in the adult, many copies of the genetic information, one (or two in diploids) in each cell. Some implications of this fact are discussed in a forthcoming book, *The*

Evolution of Individuality, by Leo Buss. For example, he thinks that the early segregation of the germ line evolved because it limits between-cell, within-individual competition for the privilege of becoming a gamete. However, I suspect that in this context the most significant thing about multicellular organisms is that, typically, they reproduce through single cells. As in the case of the stochastic corrector model discussed above, this increases the effectiveness of selection at the higher (organism) level relative to the lower (cell). This raises an interesting question concerning clonal organisms that reproduce by multicellular propagules such as bulbils or rhizomes. I can see two possibilities. One is that such organisms are doomed to death through the accumulation of selfish mutations if they do not periodically revert to reproduction through seeds or eggs. The other is that the cells of the propagule are descended from a single cell in the relatively recent past.

Demes and Social Groups

If individuals are populations of cells, and yet can evolve individual-level adaptations, can not demes, which are populations of individuals, evolve group-level adaptations? In particular, can they not evolve behavior that is favorable to group survival? This was the form of group selection proposed by Wynne-Edwards (1962). If new groups are established by single fertilized females, then genes that are advantageous to the group, but are selected against within groups, can sometimes spread. However, it is rare for higher organisms to have an appropriate population structure. It has proved to be more fruitful to analyze social behavior either from a gene-centered or an individual-centered viewpoint, allowing in either case for the genetic relatedness of interacting individuals (Hamilton 1964). In the gene-centered approach, we ask whether the phenotypic effects of a gene are such as to increase the number of copies of that gene in future generations. In the individual-centered approach, we ascribe to each individual an "inclusive fitness."

There are, however, occasions on which it seems reasonable to interpret the behavior of individuals solely in terms of the contribution they make to the success of the group, rather as one might discuss the contribution of liver

cells to individual survival. For example, Oster and Wilson (1978) discussed caste differentiation in ants in this way. The justification, I think, is that some social insects have evolved so far along the road to sociality that some individuals can ensure the transmission of their genes only by contributing to colony success. A worker ant or bee would gain nothing by leaving the colony. However, this is always a risky approach: as Trivers and Hare (1976) pointed out, workers in social insect colonies can increase their inclusive fitness by distorting the colony sex ratio, or even by producing sons parthenogenetically. It seems likely that, in all social animals, the opportunities for within-group, between-individual selection are substantial.

Species

The gene pool can be thought of as carrying the genetic information of a species, rather as the genome does of an individual. When a species splits in two, the daughter species carry most of the ancestral genes, so that "species heredity" is ensured. Species, therefore, are candidates as units of evolution. The snag, of course, is that within-species, between-individual selection is ubiquitous, and strong enough to prevent species-level adaptations.

Sexual reproduction may constitute an exception to this rule. Theory suggests that sex can confer an advantage on the population by accelerating evolution and by delaying the accumulation of deleterious mutations. The taxonomic data suggest that parthenogenetic varieties are short-lived in evolutionary time. But what of the strength of individual selection, particularly in view of the twofold cost of meiosis? One answer is that many sexual organisms have been sexual for so long, and have so many secondary adaptations associated with sex that they cannot readily abandon it. If there are no mutations to sexuality, then selection is powerless. However, there are many taxa in which closely related sexual and parthenogenetic taxa live side by side. Williams (1975) argued that the existence of such cases shows that there must be some short-term advantages to sex. It is on these cases that research should concentrate.

Even if we accept that species-level selection is important in maintaining sex, there are two reservations to be made. The first is that meiosis and

syngamy could not have evolved in the first place by such a mechanism. The second is that various processes subsidiary to sex still require a short term explanation. In particular, the rate of recombination requires such an explanation, because we know that there is heritable variation within species for the frequency and the location of chiasmata.

Groups with Cultural Inheritance

Human society is the final level of complexity so far achieved by living organisms. It depends on language, which has the consequence that information is transmitted, horizontally and between generations, by cultural rather than genetic means. The formal differences between cultural history and genetic evolution are profound, but it is intriguing that history reveals the same conflict between individual and group interests as has been fundamental in evolution.

Conclusion

In this brief account, I have omitted some levels of organization -- those of the ecological community and of the biosphere as a whole -- which are customarily included in hierarchical accounts of biological organization. This is because I do not see any sense, however stretched, in which these levels can be treated as Darwinian units of evolution. They do, however, reveal properties of stability and persistence that we find hard to explain. No Darwinist could accept the "Gaia" hypothesis, according to which the earth is analogous to a living organism, because the earth is not an entity with multiplication, variation and heredity. However, we should not be too contemptuous of that idea, logically flawed as it is, until we can give a better account of the long-term stability of the biosphere than is at present possible.

I do not think that progress is an inevitable consequence of evolution by natural selection. However, one can recognize in the evolution of life several revolutions in the way in which genetic information is organized. In each of these revolutions, there has been a conflict between selection at several levels.

The achievement of individuality at the higher level has required that the disruptive effects of selection at the lower level be suppressed. It is this common feature of these transitions, which are in other ways as disparate as that from molecule to cell and from protist to metazoan, which makes it worthwhile to compare them.

References

Buss, L. *The evolution of individuality*. Forthcoming.

Cavalier-Smith, T. 1981. The origin and early evolution of the eukaryote cell. In *Molecular and cellular aspects of microbial evolution*, ed. M. J. Carlisle, J. F. Collins and B. E. B. Moseley. Society for General Microbiology Symposium 32:33-84. Cambridge: Cambridge University Press.

Cavalier-Smith, T. 1985. *The evolution of genome size*. Chichester: Wiley.

Dawkins, R. 1976. *The selfish gene*. Oxford: Oxford University Press.

Eigen, M., and P. Schuster. 1979. *The hypercycle*. Berlin: Springer-Verlag.

Fisher, R. A. 1930. *The genetical theory of natural selection*. Oxford: Clarendon Press.

Fox, S. W. 1984. Protenoid experiments and evolutionary theory. In *Beyond neo-Darwinism*, ed. M.-W. Ho and P. T. Saunders. New York: Academic Press.

Hamilton, W. D. 1964. The genetical evolution of social behavior. *Journal of Theoretical Biology* 7:1-52.

Mayr, E. 1963. *Animal species and evolution*. Cambridge: Harvard University Press.

Oparin, A. I. 1957. *The origin of life*. New York: Academic Press.

Orgel, L. E. 1979. Selection in vitro. *Proceedings of the Royal Society, London*. B205:435-42.

Oster, G. F., and E. O. Wilson. 1978. *Caste and ecology in the social insects*. Princeton: Princeton University Press.

Szathmary, E., and L. Demeter. 1987. Group selection of early replicators and the origin of life. *Journal of Theoretical Biology 128:463-86.*

Trivers, R. L., and H. Hare. 1976. Haplodiploidy and the evolution of the social insects. *Science* 191:249-63.

Williams, G. C. 1966. *Adaptation and natural selection*. Princeton: Princeton University Press.

Williams, G. C. 1975. *Sex and evolution*. Princeton: Princeton University Press.

Wynne-Edwards, V. C. 1962. *Animal dispersion in relation to social behaviour*. Edinburgh: Oliver and Boyd.